МАТРИЧНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА ПОДБОРА ЧИСЛА КЛАСТЕРОВ РАЗМЕЧЕННОГО ГРАФА

А.П. Котенко [1,2], М.С. Щербаков [2]

[1] Самарский государственный аэрокосмический университет имени академика С.П. Королёва (национальный исследовательский университет) (СГАУ), Самара, Россия,
[2] Самарский государственный технический университет (СамГТУ), Самара, Россия

Определение числа кластеров, объединяющих вершины графа со сходными свойствами, важно в ряде задач размещения на графе оптимального числа обслуживающих центров. Оценить сходство вершин можно дать разметкой связывающих их рёбер. Приведены алгоритмы определения оптимального числа кластеров.

Ключевые слова: кластеризация, маршрутизация, метрика на графе.

Введение

Ряд прикладных задач требует разбиения заданного множества объектов на подмножества (кластеры), отражающие сходство по тому или иному признаку. Такими являются задачи размещения оптимального числа обслуживающих узлов в вершинах ориентированного или неориентированного графа: градостроительное регулирование, распределённые системы массового обслуживания, IP-маршрутизация и др.

Приемлемость и эффективность найденного разбиения в значительной степени определяется числом полученных кластеров. Как правило, требуется найти распределение подобных центров обслуживания с минимальным числом кластеров, описывающих их клиентуру.

Наиболее удобное парное сравнение объектов можно представить размещением объектов исследования в вершинах графа с рёбрами, помеченными мерой сходства-различия вершин.

Симметричность меры отношения парного сходства вершин порождает неориентированный граф. В других случаях возможно получение ориентированных графов.

Постановка задачи кластеризации

Введём следующие определения и обозначения. Рассмотрим задачу разбиения множества вершин

$$i \in \overline{1,n}, \; n := |V| < \infty ,$$

неориентированного графа $G(V,R)$ без петель с рёбрами

$i \in \overline{1,m}$, $m := |R| < \infty$, веса $|r_i| \geq 0$

на кластеры

$$i \in \overline{1,k}\ ,\ k<\infty,$$

минимизирующие расстояния $\rho(v,u_i)$ до вершин графа G, размещённых в центрах кластеров $u_i \sim U_i$, $u_i \in V$.

Центры кластеров можно считать центрами обслуживания остальных вершин. Возможна постановка задачи, учитывающая веса отдельных вершин (населённость областей жительства, интенсивность обращения клиентов из разных населённых пунктов за определённой услугой и др.). [1]

Допустим возможность пересечения отдельных кластеров (обслуживание в разных центрах), но не вложенность их друг в друга.

Такой набор кластеров $\{U_i : i \in \overline{1,k}\}$ накрывает множество вершин V, не являясь его разбиением. Поэтому допустим возможность, что число кластеров k может оказаться как не меньше, так и не больше числа n вершин графа G.

Метрику $\rho(v_i,v_j)$ между вершинами $v_i, v_j \in V$ (быть может, с потерей аксиомы отделимости) определим минимумом сумм весов рёбер пути, соединяющего данные вершины.

Отметим, что наличие аксиомы отделимости эквивалентно отсутствию рёбер нулевого веса.

Очевидно, для конечного графа G вершина $u_i \in V$, определяющая центр $u_i \sim U_i$ заданного кластера U_i, существует (возможно, не единственная) и принадлежит этому кластеру: $u_i \in U_i$.

Предложенный признак принадлежности кластерам на компонентах связности графа G будет согласован с заданной метрикой (ρ,V). Следовательно, выбор метрики ρ на вершинах графа эквивалентен выбору той или иной кластеризации множества его вершин. Изменяя определение наиболее эффективного числа кластеров, получим набор ограничений на свойства согласованных метрик, определяющих числовые характеристики отношения сходства-различия вершин графа.

Это позволяет выявить ограничения на количественные описания качественных признаков объектов, соотнесённых вершинам графа.

Постановка обратной задачи

Сформулируем обратную задачу: присвоить рёбрам $r \in R$ (не)ориентированного конечного графа G(V,R) с заданным набором кластеров числовые неотрицательные веса $|r| \ge 0$, порождающие согласованную метрику (ρ,V).

Её тривиальное решение: присвоить нулевые веса рёбрам $r := (v_i, v_j) \in R$, лежащим внутри одного кластера $v_i, v_j \in U_k$, и единичные веса – остальным рёбрам.

Тривиальное решение порождает метрику (ρ,V) без аксиомы отделимости (если хотя бы один кластер содержит более одной вершины), при которой исходные кластеры состоят из вершин, разделённых нулевыми расстояниями, а расстояние от каждой вершины до любой вершины чужого кластера больше нуля.

Возможно нетривиальное решение обратной задачи с согласованной метрикой (ρ,V), удовлетворяющей аксиоме отделимости: достаточно приписать рёбрам, лежащим внутри одного кластера, достаточно малые положительные веса, а остальным рёбрам – достаточно большие.

Точную числовую границу, разделяющую эти два типов весов, имеет конструктивное (алгоритмическое), хотя достаточно громоздкое описание.

Заключение

Кластерное представление множества вершин конечного графа эквивалентно подбору неотрицательных весов рёбер, порождающих согласованную метрику (ρ,V). Построить согласованную метрику можно, к примеру, с помощью матричного алгоритма [2].

Предложенный алгоритм обобщается на случай несимметричной меры отношения сходства-различия объектов, соотнесённых вершинам графа, переходом к орграфу с парами антиколлинеарных дуг, замещающих рёбра исходного неориентированного графа.

В таком случае получается семейство разбиений, в которых не только кластеры различны, но и их число. Строятся такие кластеры для расстояний, определённых по компонентам сильной связности при выборе по одной дуге из каждой антиколлинеарной пары.

Матричный алгоритм определения кратчайших расстояний на орграфе [2] работает и в этом случае.

Возможна также модификация алгоритма [1] для решения задачи кластеризации множества рёбер (не)ориентированного графа (дуг орграфа), вершины которого взвешены числовой характеристикой того или иного признака.

Литература

1. Kotenko, A. Labeled graphs’ vertices and edges sets clustering / Kotenko A., Bukarenko M. / Groups and Graphs, Algorithms and Automata, 2015: Abstracts of the International Conference and PhD Summer School. – Yekaterinburg: UrFU Publishing house. – 2015. – P. 41.
2. Котенко, А.П. Матричный алгоритм Беллмана–Мура // Управление организационно-экономическими системами. Самара: Изд-во СГАУ. – 2013. – Вып. 10. – С. 33-37.